

AWARD NUMBER: W81XWH-13-1-0359

TITLE: Junk DNA-Encoded Antigens in Ovarian Cancer

PRINCIPAL INVESTIGATOR: Kathleen H. Burns, M.D., Ph.D.

CONTRACTING ORGANIZATION: Johns Hopkins University School of Medicine
Baltimore, MD 21205

REPORT DATE: October 2014

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE October 2014		2. REPORT TYPE Annual		3. DATES COVERED 15 Sep 2013 - 14 Sep 2014	
4. TITLE AND SUBTITLE Junk DNA-Encoded Antigens in Ovarian Cancer				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER W81XWH-13-1-0359	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Kathleen H. Burns, M.D., Ph.D. E-Mail:kburns@jhmi.edu				5d. PROJECT NUMBER OC120390	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The Johns Hopkins University School of Medicine 733 North Broadway, Suite 117 Baltimore MD 21205				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Most endeavors to find tumor markers focus on the small fraction of the human genome that comprises unique, protein-coding exons. These exclude studies of highly repetitive DNA sequences despite the fact that this dimension of our genome is replete with protein coding potential and is known to be derepressed in cancers. In Year 1 of this pilot award, we developed an informatics approach for RNA-seq read alignment to characterize how repetitive sequences contribute to the ovarian cancer transcriptome. We also brought into our laboratory human ovarian cancer cell lines and a mouse model of ovarian cancer wherein intrabursal administration of a lentivirus expressing <i>Cre</i> recombinase inactivates homozygous floxed alleles of p53 and <i>Rbl</i> . We also increased our understanding of aberrant ORF1p long interspersed element-1 (LINE-1) protein expression in human ovarian cancer and began development of reagents to detect LINE-1 encoded ORF2p protein.					
15. SUBJECT TERMS Repetitive DNA sequences; Long Interspersed Element-1 (LINE-1); ovarian cancer					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Unclassified	18. NUMBER OF PAGES 10	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (include area code)

Table of Contents

	<u>Page</u>
1. Introduction.....	4
2. Keywords.....	4
3. Accomplishments.....	5
4. Impact.....	8
5. Changes/Problems.....	8
6. Products.....	9
7. Participants & Other Collaborating Organizations.....	9
8. Special Reporting Requirements.....	9
9. Appendices.....	10

1. INTRODUCTION:

Interspersed repeats were known colloquially as junk DNA for many years. There was little interest in their expression or functional effects. It seems a glaring omission since repetitive sequences left by transposable elements largely define the landscape of our genome. Sequences homologous to mobile elements make up more than half of our total DNA, and it is estimated that nearly two thirds of our genome has interspersed repeat content.

In humans, all recently active mobile DNAs are retrotransposons, known as ‘copy-and-paste’ transposons. These propagate through a process known as retrotransposition, which involves an RNA intermediate that is reverse transcribed to make the new insertion sequence. Each new genomic insertion is initially highly homologous to the element that templated the RNA intermediate, though this relationship between a parent element and its progeny deteriorates over evolutionary time. The repetitive nature of these sequences has posed significant challenges for researchers, for example, in assembling genome builds and for recognizing structural variants caused by recently and currently active mobile elements, although this is a rapidly advancing area.

Contributions of highly repetitive genome sequences to cellular transcriptomes have been perhaps less well understood. As sequencing methods have matured, repetitive elements remain understudied, in part because of our legacy of masking them and presuming they are nonfunctional ‘junk DNA’ as well as for more salient reasons. Some of the most significant barriers today are owed to mainstream methods for aligning next generation sequence reads. RNA-seq and chromatin immunoprecipitation (ChIP)-seq alignment algorithms typically handle the issue of ambiguously aligning reads by returning a single legitimate genomic coordinate, returning an arbitrary subset of legitimate or best alignments, or discarding the read. Showing all legitimate alignments is computationally intensive and also misrepresents the underlying biology – that there should be a correspondence between any RNA-seq read and the single genomic origin that template the RNA. Other approaches to RNA-seq alignments only consider an annotated list of gene transcripts and exclude interspersed repeats and other non-canonical RNAs.

In ovarian cancer research, as in other fields, endeavors to profile alterations in RNA expression to find tumor markers focus on the small fraction of the human genome that comprises unique, protein-coding exons. These exclude studies of highly repetitive DNA sequences despite the fact that this dimension of our genome is replete with protein coding potential and is known to be derepressed in cancers.

In Year 1 of this pilot award, we developed an informatics approach for RNA-seq read alignment to characterize how repetitive sequences contribute to the ovarian cancer transcriptome. This represents a major advance in the field. We also brought into our laboratory human ovarian cancer cell lines and a mouse model of ovarian cancer wherein intrabursal administration of a lentivirus expressing *Cre* recombinase inactivates homozygous floxed alleles of *p53* and *Rb1*. Finally, we also increased our understanding of aberrant ORF1p Long Interspersed Element-1 (LINE-1) protein expression in human ovarian cancer and began development of reagents to detect the LINE-1 encoded ORF2p protein.

2. **KEYWORDS:** repetitive DNA; interspersed repeats; long interspersed element-1 (LINE-1); ovarian cancer

3. ACCOMPLISHMENTS:

Goals of the project: Goals of the project were subdivided into four tasks:

Task 1: Characterize expression of highly repetitive sequences in human ovarian epithelial cancers.

Task 2: Describe expression of LINE-1 encoded protein in human ovarian cancers.

Task 3: Develop an assay for detecting circulating protein in human serum.

Task 4: Develop a mouse ovarian cancer model overexpressing LINE-1 tumor antigen.

Accomplishments:

(1) Development of the RepTag algorithm for next generation sequencing alignment.

The vast majority of interspersed repeats in our genome co-occurred with or antedate the activity of L1PA3/L1PA4 families of LINE-1 and are more than 10 million years old. These sequences are fixed present in genomes and shared by other primates. Some of these sequences were unique on arrival because of mistakes of reverse transcription. Others, not under the selective pressures characterized of many protein-coding exons, and have had sufficient time to accrue neutral substitutions that make each genomic location distinct from others. Still others have acquired unique junctions by being themselves interrupted by successive retroelement waves.

To identify relatively unambiguous intervals of repetitive sequences (RepTag), we:

- (i.) extracted all sequences in the reference genome assembly (hg38) annotated as a repeat by Repeatmasker as well a segment of flanking sequence 5' and 3' of the element;
- (ii.) divided these into 60-mer substrings offset from one another by one base pair (bp);
- (iii.) tested each for mappability by aligning to the entire reference genome; and
- (iv.) kept those with no legitimate matches elsewhere with 3 or fewer mismatches as unique 'tags'.

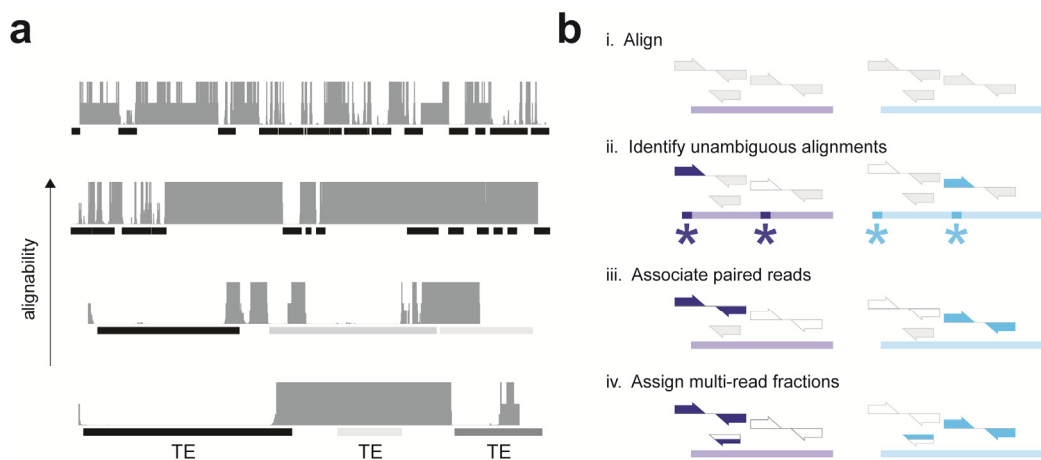


Fig 1: RepTag alignment approach. **a.** Four lengths of gDNA sequence are shown with genomic coordinates progressing left to right. The grey lines drawn upward on the Y-axis illustrate alignability of reads at each position. The bars along the lengths of the X-axis indicate the presence of interspersed repeats and their relative ages. Shading shows the homology between each interspersed repeat and its consensus sequence. Many even recently inserted repeats (black) have positions (tags) where unique sequence alignments exist. **b.** Tag positions can be used as described in the preceding paragraph to inform read assignments.

Our method for analyzing RNA sequencing reads then leverages tag alignments to identify repetitive element expression (**Figure 1**): (a.) align RNA-seq reads to the repeat library showing all legitimate alignments (BOWTIE); (b.) assign (save) alignments that correspond to tag positions and discard competing candidate alignments for these reads; (c.) assign (save) read pair alignments where these are concordant to mates assigned in (b.) and discard competing alignments; (d.) for ambiguously aligning reads, make weighted assignments between legitimate alignment positions reflecting the relative expression* levels based on reads definitively assigned in (b.) and (c.). * Here, expression reflects numbers of aligning reads normalized for the length of the repeat.

To describe the landscape of repetitive sequence expression in normal human cells, we performed RNA-seq on low passage, primary cell cultures. We prepared rRNA depleted libraries (as opposed to polyA selection), and used an Illumina library preparation that retains strand information. To accurately measure expression from repetitive elements, we used RepTag to identify uniquely aligning reads and resolve ambiguous alignments. We found that a small but distinct proportion of interspersed repeats is expressed (**Figure 2**).

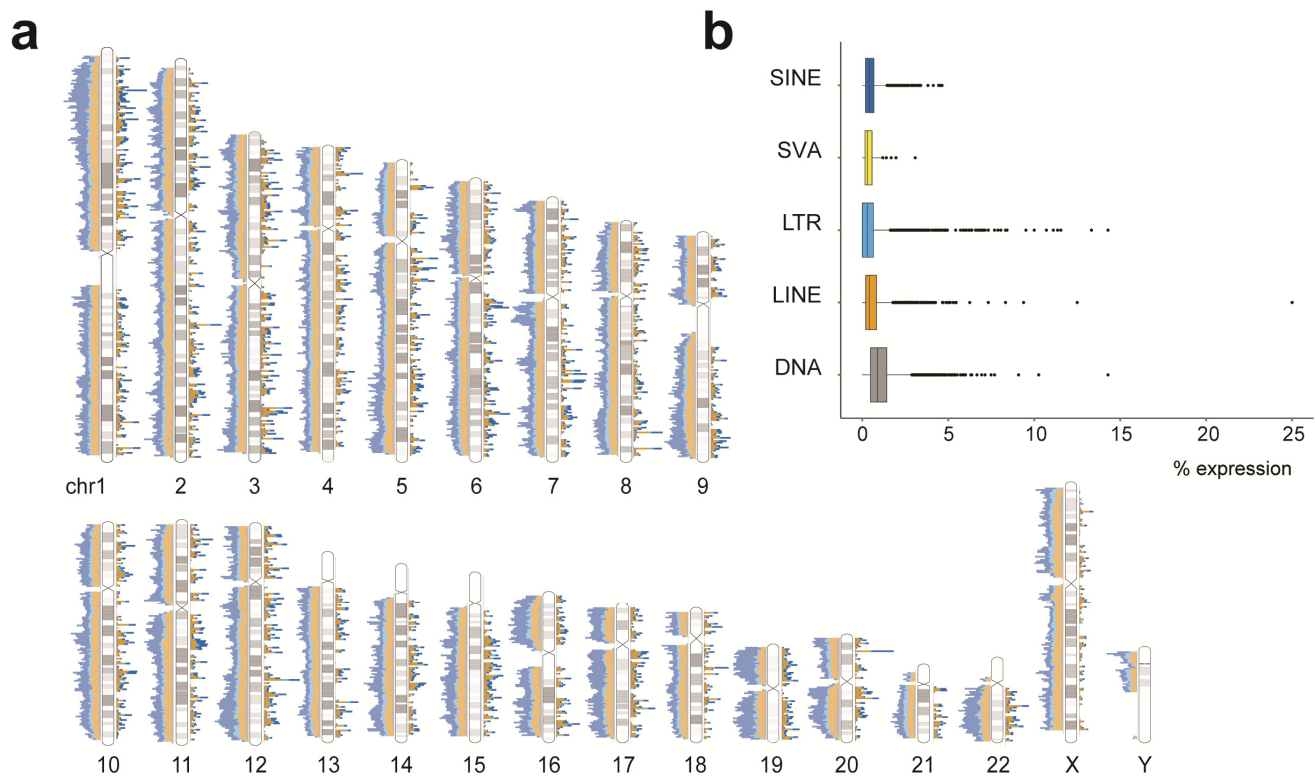


Fig 2: Interspersed repeat expression in human cells. **a.** Chromosomal ideogram showing gDNA repeat content to the left of each chromosome and RNA representation owing to interspersed repeats on the right. Interspersed repeats are color coded depending on the type of mobile DNA. **b.** Box and whiskers plots showing the proportion of each type of element that is expressed. SINE, Short Interspersed Element; SVA, SINE, VNTR (variable number tandem repeat), *Alu* composite element; LTR, long terminal repeat retrotransposons; LINE, Long Interspersed Element; DNA, DNA transposon.

(2) LINE-1 protein detection in human tumors.

During the first year of this pilot, we optimized and prepared for publication a method for immunodetection of LINE-1 ORF1p in human cancer samples using a mouse monoclonal antibody developed in our laboratory (**Figure 3**). The majority of ovarian cancers are strongly immunoreactive for LINE-1 ORF1p, an RNA binding protein encoded by this autonomous retrotransposon.

Our mouse monoclonal antibody was raised against a peptide fragment of human LINE-1 ORF1p. The reagent recognizes a sequence corresponding to amino acids 35 to 44 of human LINE-1 ORF1p (MENDFDELRE); this is in a region of relative divergence between human and mouse LINE-1 ORF1p proteins.

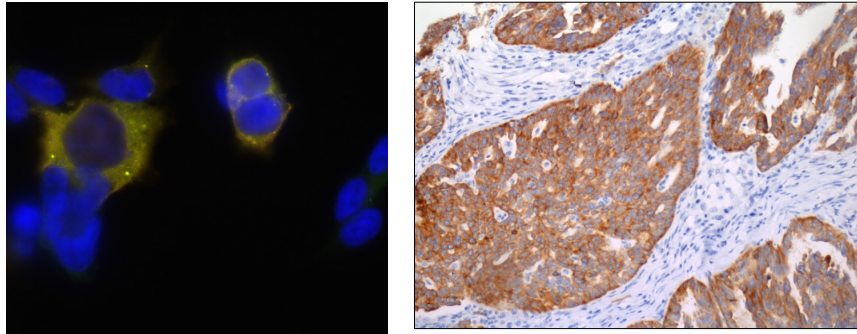


Fig 3: Immunodetection of LINE-1 ORF1p. (Left) Immunofluorescence in cells expressing a tagged protein from a transfected plasmid. Red and green channels are used to detect the FLAG tag and an epitope on ORF1p. The yellow indicates co-occurrence of this signal, which is expected. The distribution of the protein is punctate and cytoplasmic. L1 ORF1p protein expression in ovarian serous carcinoma. Polyclonal antibody (anti or α -L1ORF1p) was used. Unstained 4- μ m sections of each tissue block were kept at 65°C for 30 min prior to staining on a Bond-Leica autostainer (Leica Microsystems, Bannockburn, IL). Heat induced antigen retrieval with high pH retrieval solution was followed by a peroxide blocking step and 30 minutes of primary antibody incubation. The reaction was developed using a biotin free Bond-polymer detection (Leica Microsystems, Bannockburn, IL), and 3',3' diaminobenzidine (DAB) chromogen-substrate was used for visualization (brown). Slides were counterstained with hematoxylin (blue), dehydrated and cover slipped. Sections show an intensely immunoreactive tumor component (brown). The tumor cells infiltrate in association with a benign stromal cell component which is negative for the protein (counterstained in blue).

In Year 1 of this award, we also began development of an antibody to detect LINE-1 ORF2p. ORF2p encodes a protein with endonuclease (EN) and reverse transcriptase (RT) domains. This involved first the expression of recombinant fragments of the protein in bacterial cells and their purification (**Figure 4**).

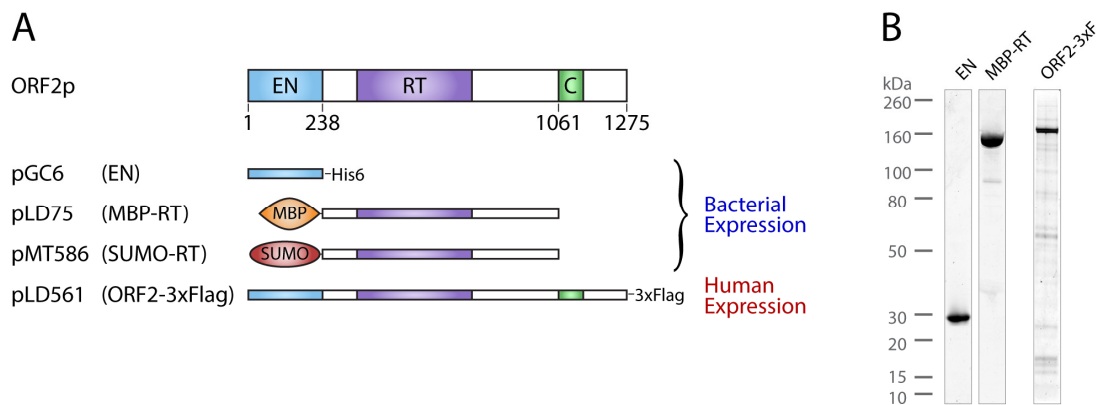


Fig 4: A. Domains of human ORF2p and bacterial and human cell expression constructs used to generate anti-ORF2 mAbs. EN, endonuclease; RT, reverse transcriptase; SUMO, small ubiquitin-like modifier. B. ORF2p domains were expressed and purified from in *E. coli* and full-length ORF2p-3xFlag expressed and purified from large-scale suspension cultures of HEK-293T cells. Coomassie blue stain. These were used to immunize rabbits for antibody production.

(3) Finally, in Year 1 of this award, we brought into our laboratory a mouse model of ovarian cancer wherein intrabursal administration of a recombinant virus expressing *Cre* recombinase inactivates homozygous floxed alleles of *p53* and *Rb1* to serve as a model of ovarian cancer development.

Briefly, this is an inducible ovarian cancer mouse model described by Flesken-Nikitin and colleagues. In their model, a single intrabursal administration of recombinant virus expressing *Cre* inactivates homozygous floxed alleles of *p53* and *Rb1*. The result is sufficient to create ovarian epithelial tumors with histopathologic similarities and similar patterns of anatomic dissemination as compared to human tumors. The *p53* and *Rb1* flox lines were obtained from the National Cancer Institute, Frederick (Stock # 01XC1 and 01XC2).

Opportunities for training and professional development provided by the project.

This project supported the academic development of Teal Scholar Wan Rou Yang. Wan Rou is an M.D., Ph.D. student whose doctoral thesis project has been supported by this mechanism. Training activities during the period of the award have included: weekly meetings with the project principal investigator to discuss project directions; presentations at group meetings for the laboratory; and poster presentations at a Department retreat.

Dissemination of results. Results will be shared in the form of scientific manuscripts.

Next reporting period. Nothing to report.

4. IMPACT:

o Impact on the development of the principal discipline(s) of the project.

We developed an informatics approach to RNA-seq data to characterize how specific genomic repeat sequences contribute to cellular RNA. *This should directly enable the identification of which of these sequences are aberrantly expressed in ovarian cancer.*

We optimized and submitted for publication a method for immunohistochemical detection of LINE-1 ORF1p in human cancer samples using a mouse monoclonal antibody we developed. *This should increase the availability and application of this cancer biomarker.*

o Impact on other disciplines.

Neither of the points above is relevant only to ovarian cancer; this project makes available informatics approaches and reagents we expect will be useful to cancer biology more generally.

o Impact on technology transfer.

Although technology transfer was not a specific goal of this project, we have worked with our institutional technology transfer office to license a mouse monoclonal antibody that we developed against LINE-1 ORF1p.

o Impact on society beyond science and technology.

This award supported research training of a student in the physician-scientist program at the Johns Hopkins University School of Medicine (JHUSOM).

5. CHANGES/PROBLEMS:

- o Changes in approach and reasons for change.** Nothing to report.
- o Actual or anticipated problems or delays.** Nothing to report.
- o Changes that had a significant impact on expenditures.** Nothing to report.
- o Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents.** Nothing to report.

- **Significant changes in use or care of human subjects.** Nothing to report.
- **Significant changes in use or care of vertebrate animals.** Nothing to report.
- **Significant changes in use of biohazards and/or select agents.** Nothing to report.

6. PRODUCTS:

- **Publications, conference papers, and presentations**

- **Journal publications.**

Yang, W.R., *et al.* Landscape of human transposable element expression. In preparation.

- **Books or other non-periodical, one-time publications.**

Sharma, R., Rodić, N., Burns, K.H., Taylor, M.S. Immuno-Detection of Human LINE-1 expression. In: Retrotransposons and Transposons. Methods in Molecular Biology. (García Pérez, J.-L., ed.) In press.

- **Other publications, conference papers, and presentations.**

Biochemistry and Molecular Genetics Seminar Series, October 2014, University of Virginia School of Medicine, Charlottesville, VA.

National Cancer Institute, Bethesda, MD. March 2015, by invitation of Kevin Howcroft, Ph.D., Chief, Division of Cancer Biology.

FASEB conference “Mobile DNA in Mammalian Genomes”, June 2015. Palm Beach, Florida.

EMBO/EMBL Symposium “The Mobile Genome: Genetic and Physiological Impacts of Transposable Elements”, September 2015. Heidelberg, Germany.

- **Website(s) or other Internet site(s)**

Nothing to report.

- **Technologies or techniques**

Nothing to report.

- **Inventions, patent applications, and/or licenses**

Nothing to report.

- **Other Products**

Nothing to report.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

- **What individuals have worked on the project?** No change.

- **Change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period.** Nothing to report.

- **Other organizations.** Nothing to report.

8. SPECIAL REPORTING REQUIREMENTS Nothing to report.

